Colleges Mine Data to Predict Dropouts

Computer analysis seeks to identify students at risk

By CATHERINE RAMPELL

When students log into the course Web site for Purdue University's freshman biology laboratory class, they see an image of a traffic light. Green means they're doing well, yellow means they're faltering, and red means they might fail.

So who's the traffic cop?

It's not the professor, and it's not the teaching assistants. Rather, it's a sophisticated computer algorithm that predicts when students are at risk of failing, based on their preparation going into the class and their behavior on the course's Web site.

Several colleges and universities like Purdue are mining data they have about students to try to improve retention. The institutions analyze years' worth of data on which students did well and which did poorly, and what variables — whether they be SAT scores, financial-aid status, or attendance at the dining halls — correlate with those successes or failures. Using those data, colleges try to predict which students are likely to drop out — and intervene before the students themselves even know they're in trouble.

At Purdue the risk algorithm is based on academic variables like GPA's and standardized-test scores, as well as how often students log into the course site. Students who have some combination of poor preparation and slack engagement with the Web site will see the red or yellow light on the course-management system and will also get a warning by e-mail asking them to meet with an instructor or seek outside help.

Purdue researchers found that students in the moderate-risk (yellow light) group who received the e-mail messages did better in the course than did their counterparts in a control group. Most of the students identified as being at highest risk (red light) still did not rectify their situations or take advantage of campus resources, however.

"It's toughest with students who are really just not engaged, who are not using the resources available at all even if you tell them to," says John P. Campbell, associate vice president of Purdue's Rosen Center for Advanced Computing, who is running the data-mining project.

"Where things get more interesting is that middle group that could slide either way," he says. "They could be B students, and they could be D students." Thanks to the early-warning system, he says, more of those students are sliding into the B group.

Secrets of Successful Students

At the State University of New York at Buffalo, a data-analysis project is tailored to engineering students. William G. Wild Jr., director of student-excellence initiatives at Buffalo's School of Engineering and Applied Sciences, has identified seven variables, including scores on standardized state math tests, that predict undergraduates' success in the highly demanding program.
Students scoring below a particular threshold on five of those variables are deemed to be at high risk of academic failure.

If they are accepted by the engineering school, those high-risk freshmen get a letter that encourages them to participate in extra-help sessions called "small groups," which are available to all engineering students. The at-risk students are not told that they have been identified as being underprepared in any way, Mr. Wild says, so that they won't feel stigmatized.

Different colleges have found different variables worth watching, and not all of them are academic. At the University System of Georgia, administrators worry about the "locus of control," a common personality-test measure the system uses to determine whether students feel they have control over their fates. At Slippery Rock University of Pennsylvania, administrators monitor how many visits students make to campus dining halls, where students have to swipe their electronic ID cards for admission. Dormitory staff members approach students who have not been to the dining hall in the first weeks of the semester to find out if they are having trouble adjusting to college life.

Some universities are looking to the students themselves for help in identifying which variables might be significant.

At the University of Alabama at Tuscaloosa, for example, students in a graduate course on data mining were asked to wade through years of raw data on incoming students and pick out factors that linked to retention using analytics software from SAS Institute Inc., a company that helped design the course.

Among their findings: Freshmen who lived off the campus were more likely to drop out.

University officials took the findings seriously and adopted a few policy changes as a result. For instance, the university began requiring first-year students to live on the campus.

Other universities have made similar decisions as a result of their data analyses. At South Texas College, a study of grade histories showed that students who enrolled late in courses frequently failed or dropped them. So, despite protests from students, officials did away with late registration.

"The data gave us the backbone to make a decision which was not very popular," says Shirley A. Reed, the college's president.

**Refining Formulas**

So far retention gains from the projects are modest — typically a few percentage points, according to officials. (Mr. Wild, at Buffalo, says the engineering school's graduation rate has increased by over a third, however.) They hope for greater gains as the new analytics-inspired policies are refined.

Of course, the predictive models are imperfect. They miss people who do not fit the typical mold of the at-risk student, and they sometimes catch students who do not need help, officials say. Rajaey Kased discovered that Buffalo's algorithm for at-risk engineering students — which he helped design and refine — would have identified him as an at-risk student, given his high-school preparation. But he says part of the reason he managed to excel — and be recruited for work on the risk algorithm — was that he voluntarily participated in the extra-help programs that the algorithm would have eventually referred him to.

Retention experts acknowledge that sometimes students who exhibit warning signs on paper overcome their statistical destinies.

"Many of the kids who are struggling have always struggled, and they know how to ask for help," says Jennifer B. Jones, director of academic retention at the University of Alabama. "Those who never struggled, and expected to get through college easily and can't, don't know what to do."

There are outliers in the other direction, too: students whose high-school records indicate they should succeed but who have done poorly or dropped out.

In the University of Central Florida's retention-focused analytics project, some of the students who are pegged as
being at highest risk of dropping out are the ones with the strongest high-school résumés, stellar SAT scores, and high GPA's, according to Ronald H. Atwell, director of the university's office of assessment and planning. "We speculate that they get to the university and they're not challenged or not motivated," he says. The university is still figuring out how to better retain those students.

'Sounds Like Big Brother'

Students' reactions to these efforts to scrutinize them and predict their futures are unclear. In most cases, the students are unaware of the efforts, and in some cases the universities try to keep the students from finding out.

"I don't care what they think of the program; I only care if it improves their grades," says Laurie E. Iten, an associate biology professor at Purdue whose class is being used to test an early-warning system based on a risk algorithm. The university declined to put The Chronicle in touch with any students in the course because the students are unaware that they are being studied.

Another student in the department, however, questioned the decision to keep most students in the dark.

"I kind of feel like this is an intrusion of privacy," says Misha R. Ownbey, a junior majoring in biological sciences and a member of Purdue's Science Student Council. "We're supposed to be adults, and this sort of sounds like Big Brother's watching. Maybe there are some people who might like it, who are too shy or intimidated to ask people for help directly. For me, I don't feel that way."

She suggests that Purdue, which is expanding its automated early-warning system to several other courses in the fall, should obtain written consent from students who wish to have their demographic and behavioral data mined.

But administrators at other universities say students appreciate the interventions, whether or not data mining is involved, because the reactions can be so personalized.

"This is kind of like in loco parentis reincarnated," says Cameron S. Cruickshank, vice president for enrollment management at Tiffin University, which assigns students who have been identified by a risk algorithm to "success coaches," or mentors, who help students manage their time and communicate with professors.

"These are not impersonal drill sergeants," he says. "It doesn't take long for students to figure out that this school really cares about them."